Report to the Center of Independent Experts on the 38th Stock Assessment and Review Committee panel (SARC 38) – Peter Shelton

Executive summary of findings

This review covers the 2003 assessment of ocean quahogs and butterfish off the east coast of the US carried out by the Northeast Fisheries Science Center. Both assessments were based on the results from delay-difference models applied to survey and catch data, incorporating information on growth rates and an assumption regarding natural mortality. Both stocks being assessed presented many difficulties. In the case of quahogs, there is very little signal in the data, but the data have little associated error. In the case of butterfish there are undoubtedly strong signals, but there is considerable error and variability in the data. The quahog biomass estimates are probably reliably scaled in absolute units because of the very good field experiment work on dredge efficiency based on a depletion model. There is very little information available to scale the butterfish estimate. Almost all of the relevant information for assessing quahogs could have been derived directly from the recent efficiency corrected surveys and yield per recruit/spawner per recruit analyses. For butterfish the assessment is marginal. A more critical review might have rejected the model fitted to the data and certainly would not have accepted the revised reference points. Simpler approaches working with relative stock size, such as the replacement index approach (Rago) may have much merit and should be investigated further in the context of this stock. The SARC requested several alternative formulations and a range of additional diagnostic output for the delay difference models applied to quahog and butterfish. These analyses were all carried out expediently by the NEFSC scientists involved in doing the assessments and were presented to the SARC in adequate form, facilitated by access to a computer network provided by the NEFSC. Given the competence and responsiveness of the NEFSC scientists involved in the assessment and the leadership of the SAW Chairman Gordon Waring, the SARC Chair, J.J. Maguire was able complete the review ahead of schedule.

Background

The SARC reviewed the assessments of two stocks: A stock of quahogs, a bivalve species, *Arctica islandica*, and a stock of butterfish, a pelagic fish species, *Peprilus triacanthus*. Quahogs in the northwest Atlantic are distributed from South Carolina north to the Scotian Shelf and the Grand Bank. The portion of the distribution being assessed was from South Carolina to Maine, considered for management purposes to be a unit stock. Butterfish in the northwest Atlantic are distributed from Florida to the Gulf of St Lawrence. The management unit is taken to be from Cape Hatteras to the Gulf of Maine.

Data in the assessments of both stocks included commercial catch data, discard data (in the case of butterfish), survey data, experimental gear efficiency data (quahogs) and a variety of biological information. The approaches taken in both assessments were similar in that a form of the Deriso-Schnute delay-difference biomass model (Quinn and Deriso 1999) was applied. The implementations involved tuning to a wide variety of data, in the genre of stock-synthesis modeling, and were programmed in C++ using AD Model Builder libraries.

Most of the emphasis in the assessments of the two stocks was on the delay difference model estimates. In quahogs some consideration was also given to efficiency corrected swept area estimates, exploitation rates obtained from the ratio of catch to efficiency corrected survey estimates and back-calculated population biomass based on recent efficiency corrected estimates and past catches. The delay-difference model for quahogs was applied separately to six subareas and provided what were thought to be acceptable fits in three of the areas. In the other three areas the emphasis was more directly on the survey and catch data. Biological reference points from SARC-27 based on yield per recruit and spawner per recruit were retained in the current assessment.

For butterfish the delay difference model was applied to the stock as a whole. The survey and catch data were very noisy and the model fits were considered marginal. Nevertheless they were used as the basis of the advisory report and were given consideration in terms of revising the reference points based on a production model fit to the surplus production estimates in the model.

Description of review activities

A number of documents were provided by NEFSC to SARC members prior to the review (see Bibliography). Not all of the documents provided had material impact on the assessments. For example an approach by Dr. Paul Rago termed the "replacement ratio method" was illustrated with respect to Gulf of Maine Haddock, and also applied to butterfish data with the purpose of informing the assessment. However the results were given little attention in the review and were not reflected in the advisory document. The method may have application in assessments where there is insufficient information to scale the absolute stock size, but nevertheless there is a need to provide advice on the status of the status of the stock and the sustainability of the fishery.

A document by Rago, Weinberg and Weidman (submitted) described the method that is used to determine the efficiency of the survey gear for quahogs, an essential component of this assessment. The quality and comprehensiveness of this manuscript was, in this reviewer's opinion, superior.

A paper on the Gulf of Maine Ocean Quahog by Schick and Porter was not presented in any detail. The working group had previously judged the data from the Maine component of the stock to be inadequate for inclusion in the assessment and the information was not reviewed by SARC

The SARC met in plenary from 1pm, Monday 17th November to 6pm and reconvened Tuesday, 18th and Wednesday, 19th November from 8.30am to 6pm on both days. During these sessions the documents listed in the Bibliography were all presented and SARC members were given adequate opportunity to question the authors, who were all present, with the exception of the authors of the paper on the quahog off Maine.

The SARC requested several alternative formulations and a range of additional diagnostic output for the delay difference models applied to quahog and butterfish. These analyses were all carried out expediently by the scientists involved in doing the assessments and presented to the SARC in adequate form, facilitated by access to a computer network provided by the NEFSC.

Given the competence and responsiveness of the NEFSC scientists involved in the assessment and the leadership of the SAW Chairman Gordon Waring, the SARC Chair, J.J. Maguire was able to move the meeting forward and was ahead of schedule in the completion of the Advisory Reports, the drafting of the SARC comments and research recommendations, and the adoption of the Consensus Summary Reports. This involved review of second, and in some cases, third drafts of certain sections of the reports. SARC accepted the revised documents and the meeting ended Wednesday at 6pm.

Summary of findings

1. Missing data in the quahog survey are currently filled in by "borrowing" values from previous or subsequent years for the same strata. It may be more appropriate to apply a GLM model with, for example, year and stratum effects. However, analyses presented to the SARC suggested that the effect of "borrowing" was minimal in the current data.
2. GLMs were applied to the commercial catch rate data (LPUE). Year, subregion and vessel were considered as possible explanatory variables. The assessment document suggested that there could be potential differences in power among fishing vessels over time. This could be tested for with a year*vessel interaction effect but this has not been done. LPUE data are not currently used as a tuning index in the quahog delay-difference model, but should they be used in the future, such investigations would be particularly important.
3. Two survey efficiency measures were examined in the quahog assessment – within year efficiency based on depletion experiments and relative efficiency between years. A bootstrap method was used to compare relative efficiency from fixed station samples in 1997 with 1999. It would also have been of interest to compare 1977 and 2002 to see if there has been any longer term change in dredge efficiency.
4. A negative relationship between dredge efficiency and quahog density was observed in the within-year experiments carried out in 2002. Similar data are presumably available for 1997 and 1999 to allow a more comprehensive analysis, but these data could not be made available to SARC at the time of the meeting. A

negative relationship between dredge efficiency and quahog density would render the survey potentially insensitive to decreasing quahog densities. This would be of serious concern. Some SARC members felt that a study of the effect of density on efficiency would require a thorough evaluation of the effects of other covariates such as depth and substrate type.

5. A comparative fishing exercise was carried out between the survey dredge and a dredge with a smaller mesh liner to determine the catchability of recruits by the survey dredge. A relative selectivity was computed based on sets in which both the survey dredge and a dredge with a smaller mesh liner caught recruit size quahogs. This analysis does not account for a potentially higher probability of zero catches in the case of the survey dredge. This could be accounted for by applying a delta model with a binomial component to account for whether or not a set is positive and a lognormal component for positive sets.
6. Efficiency of the survey dredge was computed from experimental data based on numbers of quahogs, but was applied in the assessment to correct survey biomass data. This was not considered to be a major problem because the size range of quahogs in the survey is relatively small.
7. It was suggested that LPUE for quahogs should be explored as a potential tuning index in future assessments. In at least two of the subareas LPUE trends appeared to be consistent with those determined from the survey and catch data alone, leading to greater confidence in the estimates for these regions.
8. Results were presented in the assessment from a "VPA" model for quahogs. This consisted of simply back-calculating the historic biomass values from the current efficiency corrected biomass from the surveys. The back-calculation involves adding in past catches only, under the assumption that natural mortality is balanced by growth and recruitment. This model is not fitted to the survey data, but uses the efficiency corrected survey data directly to initiate the back-calculation. It may be possible to develop this approach further, fitting to survey estimates. This would provide a model of intermediate complexity between the simple efficiency corrected survey index and the fairly complicated form of the delay-difference model.
9. Six different "scenarios" comprising decisions regarding the structure of the delay difference model and the way it was fitted to data were explored for each of the six quahog regions in the assessments. These scenarios included the exploration of the effect of estimating recruitment, assuming virgin biomass in the first year of the model run and whether or not the longer time series of survey data were used in the scaling of the absolute population. These scenarios were also compared with the last assessment and with "VPA" back-calculated estimates. In most cases there was not a lot of agreement among the scenarios or with the VPA estimates and the estimates from the last assessment, indicating considerable uncertainty in appropriate model structure for the assessment. The working group had somewhat subjectively chosen to average various "preferred" scenarios to provide a "final" estimate. In some cases the preferred scenarios included all six while in other cases this was reduced to four, based primarily on the degree to which the model fit the survey data, or was consistent with the LPUE data (not used in the fitting). SARC found this approach difficult to defend and instead the

decision was made to select the same scenario (scenario 3) for each area where the application of the delay difference model seemed to be appropriate, and for other areas to present the efficiency corrected survey estimates and the back-calculated biomass obtained from adding in the catches (i.e. the so-called VPA). Scenario 3 estimates recruitment, does not assume virgin biomass existed at the start of the modeled period, and uses only the three recent efficiency corrected survey estimates for scale while the longer timeseries of survey estimates is used for trend. In the end the delay difference model was used for three areas and the VPA for the other three areas (one of which had no catch - Georges Bank, in which case the VPA equated to simply the average of the efficiency corrected survey estimates). Total biomass over time was obtained by summing the estimates for the subareas. The application of the delay-difference model was fairly complicated and to a degree had to be treated as a "black box" by most members of SARC. It was not always clear what was actually being estimated in the model and what were the major influences on these estimates. It turned out that only three parameters actually mattered and the CV's on these parameters ranged from 54% to 91%. It seems that the application of the particular form of the delay difference model the quahog assessment may have been over-ambitious. It may have been sufficient to have carried out the simpler "VPA" computations for all areas as the basis for the assessment, relying on the efficiency corrected survey values and accurate catch data.

10. Reference points for quahogs are based on yield per recruit and spawner per recruit analyses. $F_{25\%MSP}$ is used as a proxy for F_{msy}. SARC expressed some concern that this value of F may be too high to be sustainable. This warrants further investigation.
11. Although the quahog stock as a whole was considered to not be overfished, nor was overfishing taking place, SARC expressed concern regarding the depletion that had taken place in the southern portion of the stock area.
12. A major issue in the assessment of butterfish is the estimation of discards in order to obtain an estimate of overall deaths due to fishing. Discards amounted to about 2/3 of the total catch since 1980. The working group noted that observer data are not collected according to any statistical design. This complicates any subsequent analysis. Discard information was examined from both logbook and observer data. The data were stratified by half year and landings categories and discard ratios for observer data for otter trawls were used to bump up logbook records of landings from otter trawls. Other gear types were not corrected for discards. While the stratified-bump-up approach may be adequate, it would be useful to consider a generalized linear model incorporating various explanatory variables that could be fitted to the observer data and used to predict discards in the landings data. This model would provide some diagnostics that could be evaluated.
13. For butterfish an approach called "replacement ratio" was applied (Rago). This approach looks at whether the current stock is replacing itself and whether or not the current catch is too high or too low. It appears to have some merit, particularly for data-poor stocks. It is a reduced-parameter model for non-age structured data and it does not attempt to derive absolute estimates, so the

problem of correctly scaling the survey is avoided. However, in the application to the butterfish, randomization tests suggested that there was not a lot of information in the replacement ratio-relative F relationship and the method was not pursued further in the assessment.

14. The working group had applied a delay difference model to butterfish that was similar in some respects to the model applied to quahogs. The view of several members of SARC was that this application had marginal value.
15. It was initially suggested that there was a lot of coherence in the 0 and 1+ data in the spring and fall survey, but this was disputed by SARC.
16. The delay-difference model preferred by the working group for assessing butterfish included allowing error in catches. This model appeared to predict the survey and catch data extremely closely, even though these data are considered to be highly variable, and in the case of the catch data, to be quite uncertain. This form of the model, in which catch is estimated, involves the estimation of a large number of parameters (>80). Given the quality of the available data, it seems reasonable to conclude that this model is over-parameterized.
17. The decision was made in SARC to further explore a delay-difference model for butterfish that did not include error in the catches, thus reducing the number of parameters somewhat. Even this model may be over-parameterized given the quality of the available data. This model was run with natural mortality M=0.8. Some concern was expressed that the model fit improved as M increased when a range of M was explored, although the implication of this was not immediately clear.
18. An attempt was made in the assessment to update the reference points for butterfish based on the fit of a Fox production model to the estimates of surplus production from the delay-difference model. The Fox model was chosen over the Schaefer model on the basis that the stock-recruit data appeared to be better explained using a Beverton-Holt stock-recruit model, however in reality the stock-recruit and surplus production-biomass scatters do not provide much evidence for choosing one or other model. The production model fitted the data poorly and therefore the resulting reference points are very uncertain. Stock-recruit model fits were not presented. Updating the reference points based on poor model fits may not be warranted and consideration could be given to reverting back to reference points derived from yield per recruit and spawner per recruit calculations. However, given concerns about the appropriate scaling of the butterfish biomass in the model, it is advisable that the referenced points be consistent with the model used to estimate current biomass.

Conclusions and recommendations

Both stocks being assessed presented many difficulties. In the case of quahogs, there is very little signal in the data, but the data have little associated error. In the case of butterfish there are undoubtedly strong signals, but there is considerable error and variability in the data. The quahog biomass estimates are probably reliably scaled in absolute units because of the very good field experiment work on dredge efficiency based

on a depletion model. There is very little information available to scale the butterfish estimate. Almost all of the relevant information for assessing quahogs could have been derived directly from the recent efficiency corrected surveys and yield per recruit/spawner per recruit analyses. For butterfish the assessment is marginal. A more critical review might have rejected the model fitted to the data and certainly would not have accepted the revised reference points. It may have been more appropriate to apply a production model approach rather than the delay-difference model that was attempted. However, discard and landings data might be too uncertain and survey data too variable to warrant anything but a very simple approach, probably based on relative stock size, rather than attempting to obtain absolute estimates. The working group did investigate such an approach (replacement index approach), but even this approach did not find sufficient signal in the data to warrant its use.

Recommendations:

1. The delay difference model should be fully documented and published (gray literature/internet would be adequate). It would be useful to carry out an example application with simulated data and to describe the sensitivity of the approach to assumptions and errors in the inputs. Results from the simulation study should be fully described including graphical output of likelihood surfaces and other diagnostic features such as residual plots. It would be useful to have some well thought-out reasoning regarding when to apply this approach and when not to, rather than considering it a panacea for all situations.
2. Field experiments are urgently required to determine whether or not the efficiency of the dredge used in quahog surveys is sensitive to quahog density. If the dredge becomes saturated at some intermediate level of quahog density, this could seriously compromise the assessment results.
3. The current proxy for B_{msy} in the quahog assessment should be evaluated to determine whether it is consistent with the life history of the organism.
4. More sophisticated analysis of butterfish logbook and observer data should be attempted to determine whether better estimates of discards can be obtained.
5. Assessment approaches for butterfish that are less demanding of the available data should be evaluated. It is disappointing that the replacement index approach did not work. Perhaps further consideration of how to appropriately smooth the survey data and better estimates of discarding might provide more acceptable results from this approach.

References

Quinn, T.J. and Deriso, R.B. 1999. Quantitative fish dynamics. Oxford University Press, New York. 542p.

APPENDIX I: Bibliography of all materials provided

1. Terms of Reference for both stocks
2. Agenda
3. List of working papers
4. Ocean Quahogs Assessment Document by the Invertebrate Subcommittee and Methods Working Group.
5. Description of the delay-difference (KLAMZ) model by Jacobsen.
6. A spatial model to estimate dredge efficiency by Rago, Weinberg and Weidman (submitted).
7. Gulf of Maine Ocean Quahog (*Arctica islandica*) Assessment by Schick and Porter.
8. Evidence of recent recruitment in the ocean quahog Arctica islandica in the Mid-Atlantic Bight by Powell and Mann (draft manuscript to be submitted for publication).
9. KLAMZ Modelling Methods (revised) by Jacobsen.
10. Butterfish Assessment Document by the Coastal/Pelagics Working Group and the Methods Working Group.
11. The KLAMZ (Forward Projection Model) Assessment Model (as applied to butterfish) by Jacobsen.
12. Analyses of survey precision, habitat associations and relationships to catch data for butterfish by Rago.
13. Index method and replacement ratio theory by Rago.

APPENDIX II: STATEMENT OF WORK

Consulting Agreement between the University of Miami and Dr. Peter Shelton

October 17, 2003

General

The Stock Assessment Review Committee meeting (SARC) is a formal, one-week long meeting of stock assessment experts who serve as a peer review panel for several tabled stock assessments. It is part of the overall Northeast Stock Assessment Workshop (SAW) process which also includes peer assessment development (SAW Working Groups), public presentations, and document publication within a cycle that lasts six months. The panel is made up of some 12-15 assessment scientists: 4 scientists from the NEFSC; a scientist from the Northeast Regional office, scientists from the staff of the New England and Mid-Atlantic Fishery Management Councils, and Atlantic States Marine Fisheries Commission and additional panelists from state fisheries agencies, academia (US and Canada), and other federal research institutions (US and Canada).

Designee will serve as a panelist on the 38th Stock Assessment Review Committee panel. The panel will convene at the Northeast Fisheries Science Center, Aquarium Conference Room, from November 17-20, 2003 to review assessments for Atlantic butterfish (*Peprilus triacanthus*) and ocean quahog (*Arctica islandica*).

Specific

The reviewer's duties will occupy a maximum of 14 workdays; a few days prior to the meeting for document review; the week long meeting; and a few days following the meeting to ensure that the final documents are consistent with the SARC'S recommendations and advice, and a few days to prepare the review report. No consensus opinion between two CIE reviewers will be accepted.

(1) Prior to the meeting: become familiar with the working papers produced by the SAW Working Groups (total number not final; there will be at least one per stock);

(2) During the meeting: participate, as a peer, in panel discussions on assessment validity, results, recommendations, and conclusions. Participate in the formulation of the draft SARC Advisory Report;

(3) Review the final Draft Advisory Report and Consensus Summary Report.

(4) No later than December 5, 2003, submit a written report[1] consisting of the findings, analysis, and conclusions, addressed to the "University of Miami Independent System for Peer Review," and sent to Dr. David Sampson, via email to David.Sampson@oregonstate.edu, and to Mr. Manoj Shivlani, via email to mshivlani@rsmas.miami.edu.

Contact person: Dr. Gordon T. Waring, NEFSC, Woods Hole, SAW Chairman, 508-495-2311, Gordon.Waring@noaa.gov

Signed_____ Date_____

[1] The written report will undergo an internal CIE review before it is considered final. After completion, the CIE will create a PDF version of the written report that will be submitted to NMFS and the consultant.

ANNEX I: REPORT GENERATION AND PROCEDURAL ITEMS

1. The report should be prefaced with an executive summary of findings and/or recommendations.

2. The main body of the report should consist of a background, description of review activities, summary of findings, conclusions/recommendations, and references.

3. The report should also include as separate appendices the bibliography of all materials provided and a copy of the statement of work.